Automatic Grading of Pathological Images of Prostate Using Multiwavelet Transform

Kourosh Jafari Khouzani, University of Tehran, Iran, kjafari1@yahoo.com
Hamid Soltanian-Zadeh, University of Tehran, Tehran, Iran, hszadeh@ut.ac.ir
Henry Ford Health System, Detroit, MI, USA , hamids@rad.hfh.edu

Abstract

Histological grading of pathological images is used to determine the level of malignancy of cancerous tissues. This task is done by pathologists. Pathologists are inconsistent in these judgments from day to day and from person to person. So the grades are very subjective and furthermore in some cases this is a difficult and time-consuming task. This paper presents a new method for automatic grading of pathological images of prostate based on Gleason grading system. According to Gleason grading system, each cancerous specimen is assigned one of five grades. In our method the decision is based on features extracted from the multiwavelet transform of images. Energy and entropy features are extracted from submatrices obtained in decomposition. Then a *k*-NN classifier is used to classify each image. We also used features extracted by wavelet packet and second order moments to compare various methods. Experimental results show the superiority of multiwavelet transform compared to other techniques. For multiwavelets, critically sampled preprocessing outperforms repeated row preprocessing and has less sensitivity to noise. We also found that the first level of decomposition is very sensitive to noise and thus should not be used for feature extraction.

1. Introduction

Cancer is the second killer of American people, and only cardiovascular diseases exacts a higher toll [1]. Histological grading is a very important task in the framework of prostate cancer prognosis, since it is used for treatment planning. If infection of cancer disease was not rejected by non-invasive diagnostic techniques like MRI, CT scan, and ultrasound, then biopsy specimens of the tissue is tested. For prostate, the tissue is usually stained by H&E (Hematoxyline and Eosine) technique. Then the histological grading is done by viewing the microscopic image of the tissue. This task is done by pathologists. Manual grading is very subjective due to inter- and intra-observer variations. So an automatic and repeatable technique is needed for grading. Gleason grading system is the most common method for histological grading of prostate [2]. The goal of this paper is to automate the Gleason grading.

For data classification, the decision is done based on a set of features. Since most pattern recognition tasks are first done by humans and automated later, the most fruitful source of features has been those used by the people to classify the objects. Automating the classification of objects using the same features as those used by people can be a difficult task, but fortunately the features used by machines need not be precisely those used by humans. Sometimes features that would be impossible or difficult for humans to estimate are useful in automated systems [3]. In this research, we used energy and entropy features calculated from multiwavelet coefficients of the image. Then a *k*-NN classifier was used to classify each image to appropriate grade. The leaving-one-out technique was used for error rate estimation. We also used features extracted by wavelet packet and second order moments to compare various methods. Experimental results show the superiority of multiwavelet transform compared to other techniques.

2. Gleason Grading System

There is a great need for methods to quantify the probable clinical aggressiveness of a given neoplasm, and further to express its apparent extent and spread in patients [1]. Histological grading is one of these methods. The grading of a cancer attempts to establish some estimate of its aggressiveness or level of malignancy. In Gleason grading system, the cancer may be classified as grade 1, 2, 3, 4 or 5 with increasing or lack of differentiation.

Gleason has provided a conceptual diagram in Figure 1 to show the continuum of deteriorating cancer cell architecture, and the four dividing lines along this continuum which he discovered are able to identify patients with significantly different prognosis. The Gleason system is based exclusively on the architectural pattern of the glands of the prostate tumor. It evaluates how effectively the cells of any particular cancer are able to structure themselves into glands resembling those of the normal prostate [2]. The ability of a tumor to mimic normal gland architecture is called its *differentiation*, and experience has shown that a tumor whose structure is nearly normal (well differentiated) will

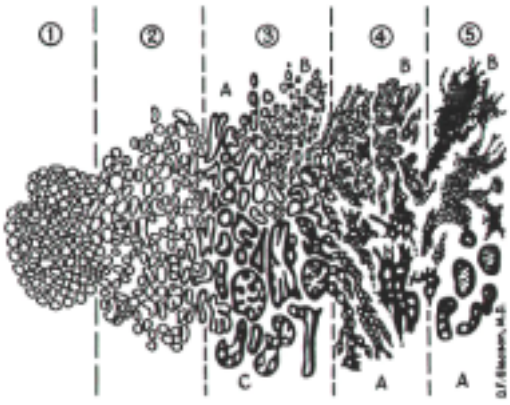


Figure 1. Gleason grading diagram.

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
25 Oct 2001	N/A	-

Title and Subtitle Automatic Grading of Pathological Images of Prostate Using Multiwavelet Transform	**Contract Number**
	Grant Number
	Program Element Number
Author(s)	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) University of Tehran Iran	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es) US Army Research, Development & Standardization Group (UK) PSC 802 Box 15 FPO AE 09499-1500	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)

Distribution/Availability Statement
Approved for public release, distribution unlimited

Supplementary Notes
Papers from 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, October 25-28, 2001, held in Istanbul, Turkey. See also ADM001351 for entire conference on cd-rom.

Abstract

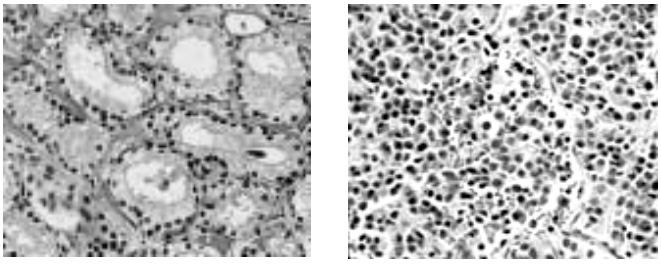
Subject Terms

Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** UU

Number of Pages
4

probably have a biological behavior relatively close to normal (that is not very aggressively malignant). Gleason grading from very well differentiated (grade 1) to very poorly differentiated (grade 5) is usually done by viewing the low magnification microscopic image of the cancer.

If there exits two patterns in the specimen, a combined score is calculated which is the sum of two grades. So combined score varies from 2 to 10. Figure 2 shows two tissue samples of grades 2 and 5. For grade 2, the glands are well-differentiated with respect to grade 5. Figure 2(b) shows only a sea of black nuclei with no pattern.



(a) (b)

Figure 2. Two samples of prostate tissue. (a) Grade 2. (b) Grade 5.

The grade of a prostate cancer specimen is very valuable to doctors in understanding how a particular case of prostate cancer can be treated. An accurate Gleason score can help one decide which treatment may be most beneficial.

In general, the time for which a patient is likely to survive following diagnosis of prostate cancer is related to the Gleason score. The lower the Gleason score, the better the patient is likely to do. Patients with score of 2 to 4 almost never develop aggressive disease, whereas most patients with a score of 8 to 10 die of prostatic carcinoma [2].

3. Feature Extraction and Classification

Wavelet decomposition has been successful in image classification and segmentation [4],[5]. The newly developed multiwavelet transform has been more successful than scalar wavelet in image denoising [6]. In this research, we used the coefficients of multiwavelet transform for feature extraction. This transform is explained next.

3.1. Multiwavelet

While in scalar wavelet transform there is only one scaling function, in multiwavelet transform we can have more than one scaling function. Multiwavelets have some advantages compared to scalar ones. For example, features such as short support, orthogonality, symmetry and vanishing moments are known to be important in signal processing. A scalar wavelet cannot possess all of these properties at the same time. On the other hand, a multiwavelet system can have all of them simultaneously. This suggests that multiwavelets may perform better in various applications [6].

In multiwavelet analysis the multiscaling function $\boldsymbol{\Phi}(t) = [\phi_1(t),...,\phi_r(t)]^{\mathrm{T}}$ satisfies a two-scale equation:

$$\boldsymbol{\Phi}(t) = \sqrt{2}\sum_k H_k \boldsymbol{\Phi}(2t-k) \qquad (1)$$

where H_k is an $r \times r$ matrix of lowpass filter coefficients. Like scalar wavelet function, multiwavelet function $\boldsymbol{\Psi}(t) = [\psi_1(t),...,\psi_r(t)]^T$ must satisfy the two-scale wavelet equation:

$$\boldsymbol{\Psi}(t) = \sqrt{2}\sum_k G_k \boldsymbol{\Phi}(2t-k) \qquad (2)$$

where G_k is an $r \times r$ matrix of highpass filter coefficients.

Corresponding to each multiwavelet system is a matrix-valued multirate filterbank, or multifilter shown in Figure 3. The lowpass filter and highpass filter consist of coefficients corresponding to the dilation equation (1) and wavelet equation (2) and these coefficients are matrices, so during the convolution step they must multiply vectors (instead of scalars). This means that multifilter banks need input rows. Thus, some methods for vectorization of scalar input should be used. Methods for preprocessing have been developed [7],[8]. In this research, we used repeated row and critically sampled approaches.

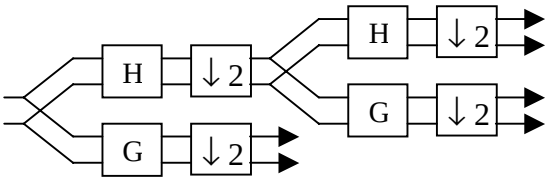


Figure 3. Multirate filterbank, showing 2 levels of decomposition.

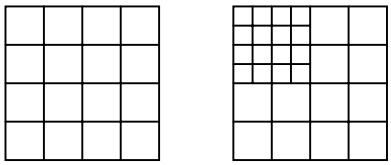
3.2. Multiwavelet Transform of 2-D signals

For calculating multiwavelet transform of 2-D signals, we can use tensor product method, i.e., performing the 1-D algorithm in each dimension separately [6]. Figure 4 shows the submatrices resulted from 2-D multiwavelet decomposition. The result after one decomposition can be realized as the following matrix (Figure 4(a)):

$$\begin{matrix} L_1L_1 & L_2L_1 & H_1L_1 & H_2L_1 \\ L_1L_2 & L_2L_2 & H_1L_2 & H_2L_2 \\ L_1H_1 & L_2H_1 & H_1H_1 & H_2H_1 \\ L_1H_2 & L_2H_2 & H_1H_2 & H_2H_2 \end{matrix}$$

Here the subband labeled L_1H_2 corresponds to data from the second channel highpass filter in the horizontal direction and the first channel lowpass filter in the vertical direction. The next step of decomposition will decompose the following "low-lowpass" submatrix, in a similar manner:

$$\begin{matrix} L_1L_1 & L_2L_1 \\ L_1L_2 & L_2L_2 \end{matrix}$$



(a) (b)

Figure 4. Result of 2-D multiwavelet decomposition. (a) One level of decomposition. (b) Two levels of decompostion.

This is shown in Figure 4(b). The number of submatrices will be equal to 4+12*l* where *l* is the number of levels of decomposition.

3.3 Feature Extraction

The features used for classification are calculated from energy and entropy of the multiwavelet coefficients. As indicated in Section 3.2, the result of decomposition is a number of submatrices. From each submatrix $[x_{ij}]$, the following features are calculated:

$$Energy = \frac{\sum_i \sum_j {x_{ij}}^2}{N \times N} \quad (3)$$

$$Entropy = \frac{-1}{\log N^2} \sum_i \sum_j \left[\frac{{x_{ij}}^2}{norm^2} \right] \log \left[\frac{{x_{ij}}^2}{norm^2} \right] \quad (4)$$

where $norm^2 = \sum_i \sum_j {x_{ij}}^2$ and N is the dimension of each submatrix.

3.4. Classification

Having vector of features of data set of images, a *k*-nearest neighbors (*k*-NN) classifier using Euclidean distance was used for classification. Because of limit to the size of the data set, we used the leaving-one-out technique to estimate error rate. Before classification, we did some normalization on the features. Recall that if one of the features has a very wide range of possible values compared to the other features, it will be a very large effect on the total dissimilarity, and the decisions will be based primarily upon this single feature. To overcome this, it is necessary to apply scale factors to the features before computing the distances [3]. In this research, we normalized each feature to have mean of zero and standard deviation of one for the entire data set. Furthermore, because some features may be more important than others, we used weight for each normalized feature. To calculate the best weight vector for the feature vector, we minimized the error rate estimated by the leaving-one-out technique.

Two different criteria were assumed to evaluate the error. The first criterion assumed the result of each classification true, if in *k* neighbors, more than *k*/2 images were of the same class. We used this criterion in error minimization (Table 1). During error minimization this criterion leads to more separation of classes in feature space. The second criterion assumed the classification result true, if most of the *k* neighbors were of the same class. The results in Table 2 are based on this criterion. It is obvious that the first criterion leads to more error (Because some images are misclassified). Furthermore as we will see in Table 1, this criterion leads to higher error for even *k*'s. For example for *k*=2 and *k*=3 this criterion means that for true classification of an image of class *c*, at least 2 of its *k* neighbors should be of class *c*. So, this leads to more error for *k*=2 compared to *k*=3.

3.5 Noise Effect

We also used a set of images as a test set and evaluate the noise effect by adding Gaussian noise with SNR=10 to images before classification, where SNR is the ratio of signal energy to noise energy. We used the second criterion in section 3.5 for error evaluation.

4. Experimental Results and Conclusions

In our experiments, 100 graded prostate tissue sample images were processed by the proposed approach. These images were of grades 2 to 5 and of magnification 100. Grade 1 was excluded because it is a very rare pattern and should be avoided.

We first made each image black and white, then decomposed it to submatrices. A set of features using (3) and (4) was calculated, and then normalized. First and second levels of decomposition were tested using GHM [9], CL [10] and SA4 [11] multiwavelets. The *k*-NN classifier was tested for *k*=1,2,…,5. Furthermore, for comparison, other features using wavelet packet decomposition and second order moments as defined in [12] were calculated. We used Daubechies wavelet D_6 and D_{20} for wavelet packet that had better results compared to other Daubechies wavelets. Table 1 shows the estimated errors using leaving-one-out technique. In this table, r.r. and c.s. show repeated row and critically sampled preprocessing respectively. The results show the superiority of multiwavelet transform for grading, with respect to other techniques.

We also divided our set of images randomly into two 50 images groups, and used one set as reference set and the other as test set. Then Guassian noise with mean zero and SNR=10, was added to test images. The results of classification of noisy data are in Table 2. These results are the average of error for 10 realization of Guassian noise. The results are rounded.

We can see that the first level of decomposition is very sensitive to noise. So we should ignore this level. This helps to noise reduction. Also for second level, critically sampled preprocessing has lower sensitivity to noise compared to repeated row preprocessing. This is due to compact form that critically sampled technique can produce. This leads to higher energy and so higher SNR at low resolutions and so less sensitivity to noise.

For next researches, better classification can be reached using the combination of second and higher levels of decompositions. But these levels may have common information. So we should select best features for classification (One of the drawbacks of multiwavelets in feature extraction is the large number of produced features).

Table 1. Percentage of error rates using multiwavelet, wavelet packet and second order moments.

k			1	2	3	4	5
1st level of decomposition	GHM	r.r.	28	35	21	27	26
		c.s.	6	18	11	20	16
	CL	r.r.	12	27	20	31	25
		c.s.	6	17	9	15	13
	SA4	r.r.	11	29	17	30	23
		c.s.	7	15	11	19	14
2nd level of decomposition	GHM	r.r.	14	28	18	34	19
		c.s.	8	17	10	20	17
	CL	r.r.	8	21	11	20	15
		c.s.	6	17	10	21	18
	SA4	r.r.	12	32	18	33	24
		c.s.	6	15	8	17	14
Wavelet Packet D_6			14	30	16	34	25
Wavelet Packet D_{20}			13	28	17	34	28
2nd order Moments			19	38	27	39	37

Table 2. Percentage of error rates for noisy data.

k			1	2	3	4	5
1st level of decomposition	GHM	r.r.	42	48	43	41	39
		c.s.	38	45	45	45	49
	CL	r.r.	38	41	44	43	39
		c.s.	38	45	45	46	47
	SA4	r.r.	36	43	43	44	41
		c.s.	39	36	36	35	41
2nd level of decomposition	GHM	r.r.	42	43	46	41	46
		c.s.	16	24	26	28	30
	CL	r.r.	34	35	34	37	30
		c.s.	20	16	24	24	32
	SA4	r.r.	36	34	35	34	35
		c.s.	16	22	24	20	28

References

[1] Kumar, Cotran, Robbins, *Basic Pathology*, sixth edition 1997, W.B. saunders company.

[2] J. Rosai, L.V. Ackerman, *Ackerman's Surgical Pathology*, 8th edition, Mosby Inc., 1996.

[3] E. Gose, R. Johnsonbaugh, and S. Jost, *Pattern Recognition and Image Analysis*, Prentice Hall, 1996.

[4] A. Laine and J. Fan, "Texture classification by wavelet packet signatures", *IEEE Trans. Pattern Analysis and Mach. Intell.*, vol. 15, no. 11, pp. 1186–1191, 1993.

[5] C. Lu, P.C. Chung and Chih Chen, "Unsupervised texture segmentation via wavelet transform," *Pattern Recognition*, vol. 30, no. 5, pp. 729-742, 1997.

[6] V. Strela, P. Heller, G. Strang, P. Topiwala, and C. Heil, "The application of multiwavelet filter banks to signal and image processing," *IEEE Trans. Image Processing*, vol. 8, no. 4, pp. 548-563, 1999.

[7] X.G. Xia, J.S. Geronimo, D.P. Hardin, and B.W. Suter, "Design of prefilters for discrete multiwavelet transforms," *IEEE Trans. Signal Processing*, vol. 44, pp. 25-35, 1996.

[8] D.P. Hardin and D.W. Roach, "Multiwavelet prefilters I: Orthogonal prefilters preserving approximation order p<=2," *IEEE Trans. Circuits and Systems*, vol. 45, no. 8, pp. 1106-1112, Aug. 1998.

[9] J.S. Geronimo, D.P. Hardin, and P.R. Massopust, "Fractal functions and wavelet expansions based on several functions," *J. Approx. Theory*, vol. 78, no. 3, pp. 373-401, 1994.

[10] C.K. Chui and J.A. Lian, "A study of orthonormal multiwavelets," *Appl. Numer. Math.*, vol. 20, pp. 273-298, 1995.

[11] L.-X. Shen, H.H. Tan, and J.Y. Tham, "Symmetric-antisymmetric orthonormal multiwavelets and related scalar wavelets," *Applied and Computational Harmonic Analysis (ACHA)*, vol. 8, no. 3, pp. 258-279, May 2000.

[12] A.P. Dhawan, Y. Chitre, C. Kaiser-Bonasso and M. Moskowitz, "Analysis of mammographic microclcifications using gray-level image structure features," *IEEE Trans. Medical Imaging*, vol. 15, no. 3, pp. 246-259, 1996.